

Títol: Sistema Recomanador d'events en una xarxa social

Autor: Maria del Mar Llambí Morillas u1915286

El projecte consisteix en construir un sistema recomanador que basant-se en l'activitat dels usuaris en l'aplicació, recomani un llistat d'esdeveniments a aquest. El projecte s'engloba dins de l'àmbit de la intel·ligència artificial, més concretament els recomanadors col·laboratius que explicarem més endavant. Ens basarem en diferents maneres d'enfocar les dades per obtenir esdeveniments amb procediments diferents. Per aconseguir-ho es divideix aquest projecte en tres pilars:

- El desenvolupament d'una aplicació web per poder interaccionar amb l'usuari i no només recollir les dades necessàries d'aquest, sinó mostrar els resultats i recollir el feedback. Aquesta permetrà registrar-se, loguejar-se, establir amistats amb altres usuaris, buscar usuaris, buscar esdeveniments, filtrar esdeveniments, mostrar esdeveniments i establir el vincle entre l'usuari i l'esdeveniment.
- La implementació d'un algorisme que a partir de les dades recollides en l'aplicació, ens calculi aquestes recomanacions. Aquest algorisme tindrà diferents bases que es descriuen més endavant i a les quals es vol assolir en totalitat. Basant-nos en el sistema col·laboratiu, s'implementarà paral·lelament un model *Nearest Neighborhood* i un model *All Neighborhood* i s'aniran complementant amb nous filtres sorgits de la metodologia escollida per desenvolupar el projecte.
- La recollida del *feedback* dels usuaris després de l'ús de l'aplicació amb el sistema recomanador i l'elaboració d'unes estadístiques de funcionament com a conclusions del projecte.

1- Conceptes prèvis

a. Sistema recomanador

La construcció de sistemes que acompanyen a l'usuari en la seva presa de decisions és una de les metes dels sistemes recomanadors. Sobretot, en una recomanació de qualitat buscant en primer lloc l'eficiència.

El repte de proveir una recomanació assequible, personal i de màxima qualitat és l'eix central dels dos pilars d'aquest sistema que els dividirem entre els aspectes tècnics i els aspectes psicològics. És a dir, parlem de tècnics quan l'objectiu és trobar regles i mètodes que explorin i explotin la informació existent d'una manera efectiva i eficient;

els factors psicològics són tots aquells que acompanyen les dades a l'hora de presentar-se davant de l'usuari i els factors que engloben a l'usuari en el moment d'entrar dins el sistema (estat anímic, caràcter, circumstàncies personals...). Proveir un portal de recomanacions personalitzades però, implica que aquest sistema ha de saber-ho tot sobre l'usuari. Introduïm doncs una primera idea que seria el perfil d'usuari o el model usuari que genera un sistema recomanador per a poder treballar sobre aquest usuari. No cal que sigui complex, per començar podria tenir per exemple les preferències i gustos de la persona. La construcció d'aquest model és clau per al sistema recomanador i la manera de construir-lo depèn de cada aplicatiu. Per a construir un bon model d'usuari ens els aplicatius actuals però, ens falta un últim pilar: la comunitat. Aquest entorn comunitari ens proporcionarà informació addicional com les amistats del nostre usuari i per tant, les preferències d'aquestes amistats. Construirem doncs el nomenat *neighborhood* o mostra de veïns que podrà anar des de tot el conjunt d'usuaris fins a subgrups que realitzi el propi usuari de les seves amistats.

b. Sistema col·laboratiu

La idea principal de l'elaboració d'una recomanació col·laborativa és explotar la informació sobre el comportament anterior o les opinions dels usuaris existents en la comunitat per a predir quins ítems o productes probablement més li agradarà o no a l'usuari actual.

Per enfocar-ho i fer-nos una primera idea de com funciona, imaginem-nos una matriu d'usuaris i ítems i com a valors, per exemple, les seves avaluacions o ràtings sobre els productes que volem estudiar.

Volem saber per un usuari quina serà la predicció d'avaluació o rating segons les reaccions de la resta d'usuaris i les seves en esdeveniments anteriors. Per això aplicarem la fórmula:

$$pred(a,p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a,b)}$$

Per calcular la similitud entre l'usuari a i qualsevol usuari de la matriu (b) utilitzarem la Correlació de Pearson.

c. Filtratge per contingut

No és tan fàcil trobar característiques de l'usuari que sempre ens serveixin: perquè un cop t'hagi agradat un llibre no vol dir que tots els llibres que pertanyen al mateix

gènere t'agradi, o perquè el llibre fos de tapa tova no vol dir que se t'hagi de recomanar tots els llibres de tapa tova.

Els processos més avançats d'intel·ligència artificial són capaços de generar similituds entre ítems de manera bastant precisa; un exemple seria el *More Like This* que implementa *ElasticSearch*. En el cas d'aquest projecte, s'ha utilitzat el contingut per aplicar un filtre després del sistema col·laboratiu. Un cop s'obtenen les prediccions de recomanació, es passa un filtre basat en contingut per augmentar o disminuir el valor de la probabilitat $p(a,r)$ on a és l'usuari i r l'ítem.

2- Tecnologies utilitzades

- IDE Eclipse
- Repositori Bitbucket i control de versions Git
- Framework CakePHP per a l'aplicació web (amb PHP, Javascript i HTML)
- Bootstrap template per una web responsive
- Servidor en Cloud Amazon Web Service
- Base de Dades MySQL
- Java i llibreria Apache Maths per al Sistema Recomanador
- Board Trello for Agile development

3- Conclusions de la implementació

- **Ampliar el coneixement sobre Sistemes Recomanadors**

Aquest projecte m'ha proporcionat la oportunitat d'endinsar-me en el la branca de la intel·ligència artificial que més interès em despertava. He estat capaç de realitzar un curs sencer *online* sobre introducció a aquests sistemes i assolir els coneixements bàsics per planificar-ne un projecte.

- **Aprenentatge de noves tecnologies i metodologies**

Durant aquest projecte no només s'ha aconseguit més coneixements teòrics sinó que també s'ha millorat les habilitats en molts aspectes tècnics.

- Millorar el coneixement en Java
- Ampliar el coneixement de CakePHP i entendre Bootstrap
- Ampliar coneixement de Javascript

- Funcionament i configuració de AWS

- **Entendre l'ètica sobre la captura de dades implícites de l'usuari**

El gran repte que s'obra dins aquest món és el de ser capaços de crear sistemes automatitzats i auto-aprenents que tinguin uns límits en quant a la recol·lecció de les dades es tracta. “Amb l'ètica hem topat” diuen els grans experts en aquesta àrea i és que, si algú pensava que la única limitació és la purament científica, és que no ha entès l'essència dels sistemes recomanadors. Ha estat curiós entendre aquest món i consultar la legislació europea que es vol fer al respecte.

- **Les limitacions dels sistemes recomanadors**

Entre la gran investigació que hi ha d'aquest món, és poca si tenim en compte la gran quantitat de limitacions que té. L'inici en fred, la dispersió de dades, la introducció d'un nou element a avaluar en una gran base de dades, el creixement de les dades, l'ètica regulada en la legislació vigent, la sinonímia, diversitat d'opinions i la identificació d'usuaris diferents.

- **El desenvolupament d'una xarxa social per poder interaccionar amb l'usuari**

S'ha aconseguit l'objectiu de desenvolupar un aplicatiu web amb totes les característiques de xarxa social per recopilar informació i construir així una interfície on es mostri també l'*output* del sistema.

- **La implementació d'un Sistema recomanador a partir de les dades escollides**

Aquest projecte és, sense cap mena de dubte, el primer esgraó en aquesta especialitat i el que m'ha fet obrir els ulls a tot una disciplina nova. La importància principal d'aquest projecte no era fer un sistema recomanador perfecte, que si era important però no principal, sinó assentar les bases d'un coneixement per a un treball futur amb un projecte des d'on partir.

- **La recol·lecció de les opinions dels usuaris**

Les proves del funcionament d'un aplicatiu es fan a partir de test automàtics i test d'usuaris. Així, se'n verifica el seu correcte funcionament i es detecta els possibles errors que puguin haver-hi. L'el sistema recomanador? es tractava d'un software que només el podien validar aquells usuaris els quals se'ls hi havien recopilat les dades. Per suposat l'avaluació és subjectiva però, com he pogut comprovar en l'anàlisi de les dades, ajuda al programador a fer-se una opinió d'allò que desenvolupa.

- **La implementació d'un sistema que es nodreixi d' Elasticsearch**

Aquest era un dels objectius inicials que es volia desenvolupar en aquest projecte. Si bé s'ha explicat en apartats anteriors que la seva raó de ser-hi perdia força per el canvi d'enfocament que s'havia donat al treball, tampoc hauria estat descabellat incorporar-lo. El filtratge per contingut hauria estat més complex encara que no hagués funcionat 100% amb tots els esdeveniments. L'altre motiu de no incloure-ho, que s'explica en aquestes línies, va ser un motiu personal. Per indisposició mèdica vaig perdre el temps necessari per desenvolupar-ho i incloure-ho al projecte i després de fer una valoració de la importància del *backlog*, es va descartar.

- **Planificar i desenvolupar un projecte informàtic**

Se'ns explica a la carrera que, un bon enginyer no és aquell que sap desenvolupar millor que un altre les habilitats tècniques (com podrien ser programació), sinó aquell que és capaç de tenir els coneixements suficients per planificar un projecte segons els seus requeriments. Què significa això? identificar quines tasques necessitaran més temps que altres no per la quantitat de feina, sinó per la complexitat, resoldre conflictes de calendari quan quedin tasques per fer i no s'arribi als terminis, ser capaç de redactar una memòria on quedin reflectits tots els mesos de treball...

4- Conclusions dels resultats

Com s'ha comentat en apartat anteriors, els possibles errors de l'aplicatiu *web* han estat corregits gràcies al *feedback* dels usuaris. Però, com verifiquem que el sistema recomanador funciona? Així doncs, es va demanar a tots els usuaris que s'havien registrat a la pàgina web que emplenessin un formulari on havien de donar tres puntuacions: la qualitat dels esdeveniments dels apartats recomanats, top 20 i *random*.

A partir dels resultats es realitza un diagrama de caixa, histogrames i es planteja una hipòtesi sobre les mitjanes de les tres valoracions en una població i es demostra a partir del test de mitjanes contrastades. És amb aquestes dades que s'acaba provant la qualitat final del sistema recomanador i com compleix l'objectiu inicial marcat.